# PRE-VALIDATION OF SERUM/PLASMA BIOMARKERS FOR LUNG CANCER

**Table of Contents:** **Page**

# PREVALIDATION OF SERUM/PLASMA BIOMARKERS FOR LUNG CANCER

## 1. Study Goals:

The initial goal of the NCI/EDRN/SPORE Lung Cancer Proteomics Committee (LCPC) is to develop the requisite sample resources to pre-validate serum/plasma biomarkers for the early diagnosis of lung cancer. Researchers may use these resources and process for continued biomarker discovery but this is not the primary activity of the Committee.

Our specific goals include:

1. Develop reference case/control serum/plasma sets held in a NCI repository and make these samples available to the research community.
2. Define, refine and validate blood-based biomarkers for lung cancer.
3. Test reproducibility of biomarkers within and across institutions.
4. Test reproducibility of biomarkers within and across analytical platforms.

## 2. Background:

A series of serum/plasma biomarkers proposed to be ready for validation were presented at the September 19, 2004 NCI Early Detection Research Network (EDRN) Lung/Upper Aerodigestive Tract Collaborative Group meeting in Denver. Several groups in the EDRN, the Lung Cancer SPOREs, and others have established candidate biomarkers for lung cancer and are in the developmental pipeline and many with promising results. We identified a *need for one or more reference sets of blood samples, serum or plasma (and tissue samples in the future), to evaluate promising biomarkers for early detection of lung cancer*. There was consensus that markers need to be tested in samples from different institutions and analytical approaches need to be tested in different laboratories.

### *a. Biomarker evaluation*:

Evaluation is needed in at least 3 specific contexts:

**1.** *Rapid single biomarker pre-validation set:* A biomarker with demonstrated good performance in a discovery sample set, typically from one institution, needs to be tested in samples from *independent clinical populations* before going forward to validation in the combined pre-validation set or in a multi-center validation study.

**2.** *Panel of biomarkers pre-validation set:* The performance of an individual biomarker is insufficient alone but may have utility when *combined in a panel* with other biomarkers measured in the same sample set. In addition to this use, the combined pre-validation set may also be used in lieu of a multi-center validation study to validate a single marker or to construct and validate a pattern analysis marker (e.g., a proteomics profile), which can be considered to be a single marker with a complex decision rule.

This combined pre-validation set must satisfy two purposes, and hence will have two sets of specimens. In this evaluation, investigators will establish (in the "construction set") a combination marker decision rule(s). In the second set of specimens (the "test" set, see **3**.) they will evaluate the constructed rule(s) in an independent sample of specimens to determine the rule's operating characteristics (e.g., sensitivity, specificity, positive and negative predictive value). Since all researchers utilizing the combined pre-validation set will be evaluating their markers on essentially identical sets of specimens and contributing

the data from their assays to the central database of assay results, over time the reference set will become progressively more valuable as the number of possible marker combinations that may be constructed and tested using this reference set expand.

**3.** *Phase II biomarker validation set:* for single or combined biomarkers which have passed **1.** or **2.**

***b. Controls*:**
Members of the Committee agree that controls should include blood samples from groups of individuals with matched age, sex, race, smoking status and smoking pack year history of smoking to the lung cancer cases. In addition, there is consensus that these controls include samples from subjects with a clinically relevant spectrum non-malignant lung disease and malignant disease from other organ sites.

***c. Sample collection and standardization*:**
A critical aspect of serum/plasma biomarker validation is to be able to confirm the predictive value of biomarkers beyond the heterogeneity in sample collection, preparation, storage and shipping (*1*,*2*). While difficult to achieve in retrospective sets, the Committee agreed to focus on initially obtaining retrospective samples from institutions where cases and controls were collected, processed, and stored under the same, or closely related, protocol.

**3. <u>Rationale for the two reference sample sets:</u>**

**REFERENCE SET A** will focus on pre-validation of biomarkers of *diagnosis of lung cancer* and target lung cancer diagnosed for individuals at high risk for lung cancer or abnormal chest x-ray (CXR) or chest computer tomography (CT) but outside of the context of a CT screening trial. **The clinical question** to be tested after pre-validation relates to whether a serum/plasma biomarker has added value to current clinical tests (CT scan and/or PET scan) for the diagnostic evaluation of pulmonary nodules and to whether such a biomarker could reduce the number, and the attendant cost, of unnecessary invasive tests (PET or tissue biopsy) or futile thoracotomies.

**REFERENCE SET B** will focus on pre-validation of biomarkers of *early diagnosis (screening) of lung cancer* and targeting a specific population of lung cancer patients diagnosed in the context of a computed tomography (CT)-based screening trial of high risk individuals. **The clinical question** to be tested after pre-validation relates to whether a serum/plasma biomarker has added diagnostic value to current tests (CT scan and/or PET scan) for the diagnostic evaluation of CT-detected pulmonary nodules.

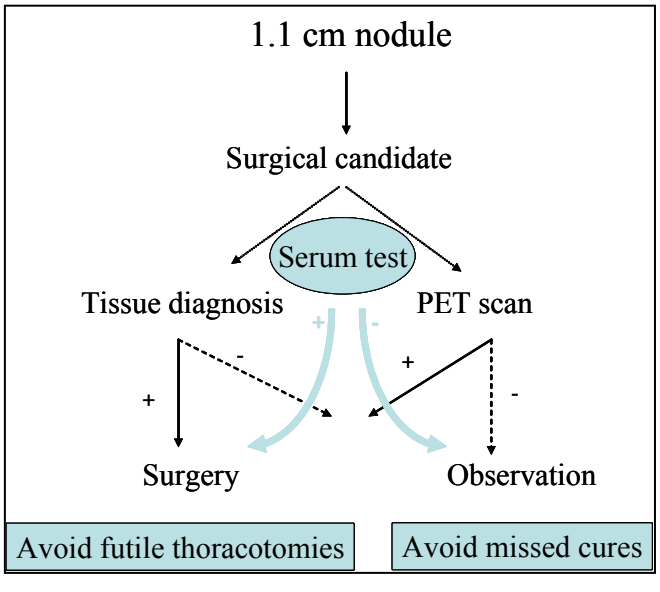



**Figure 1. Schematic flowchart illustrating one of the potential uses of a serum/plasma test in the diagnostic approach to lung cancer.** Suppose a hypothetical suspicious non-calcified, peripheral nodule of 1.1 cm in diameter is found fortuitously on chest CT. Should the patient be a surgical candidate, the management of this nodule may suggest two clinical options. A PET scan can be obtained and, if found negative, may lead to observation and repeat CT at 3-6 months, accepting the risk of missing the opportunity for early surgical intervention and cure should it really be a malignant lesion. Should the PET scan be positive, the physician may recommend resection. Should the physician decide to obtain a tissue diagnosis (i.e. fine needle aspirate or biopsy), this invasive approach, if confirmatory for cancer, will lead to surgery, and, if negative, may influence the patient and the physician to either observe the lesion accepting the risk of missing a cure, or to undergo surgery and accepting the risk of a futile thoracotomy.

## 4. Description of reference serum sets:

### REFERENCE SET A:

1. Clinical setting: Diagnosis of lung cancer

2. Cases eligibility:

- All lung cancers discovered on CXR or on CT .
- Pathology: confirmation of malignancy, all histological groups, primary lung cancer
- ≥50% Stage I
- No prior history of lung and other cancer (except for basal cell carcinoma of the skin) in the last 5 years

3. Controls eligibility
- High risk individuals as defined by ≥50 YO, ≥ 30 PKYs of smoking (prevalence of cancer ~1% based on (*3-7*)).
- Individuals with lung lesions on CXR or on CT suspicious for lung cancer but proven not to be cancer at 1 year follow up.
- 75 patients with pathology proven primary cancers from other organ sites (25 breast, 25 colon, 25 prostate)

Cases and controls will be matched for prevalence according to age, sex, race, smoking status, and PKY history of smoking

4. Sample sizes

| | |
|---|---|
| (1) Rapid single biomarker pre-validation: | 87 cases and 50 controls, 25 other cancer controls |
| (2) Panel of biomarkers pre-validation: | 150 cases and 150 controls |
| (3) Phase II validation: | 180 cases and 180 controls, 75 other cancer controls |

5. Institution provider candidates

Pittsburgh
Vanderbilt
MDACC
Moffitt
UCLA
UCHSC
NYU
JHU

**REFERENCE SET B:**

1. Clinical setting

   CT screening trial for the early detection of lung cancer

2. Cases eligibility:

   - Detected by CT screening
   - No prior history of lung or other cancer except for basal cell carcinoma of the skin in the last 5 years
   - Size: Lung cancer $\geq$0.5cm and $\leq$3cm
   - Pathology: confirmation of malignancy, all histological groups, primary lung cancer

3. Controls eligibility:

   - High risk individuals as defined by $\geq$50 YO, $\geq$ 30 PKYs of smoking (prevalence of cancer ~1%) undergoing screening chest CT
   - CT with a lung nodule $\geq$0.5cm and $\leq$3cm, free of disease at the 1 year F/U CT
   - CT without a lung nodule, free of disease at the 1 year F/U CT
   - 75 patients with other cancers (25 breast, 25 colon, 25 prostate)

   Match prevalence according to age, sex, race, smoking status, and PKY history of smoking

4. Sample size

   (1) Rapid single biomarker pre-validation: 38 cases and 87 controls, 25 other cancer controls
   (2) Panel of biomarkers pre-validation: 150 cases and 150 controls
   (3) Phase II validation: 170 cases and250 controls, 75 other cancer controls

5. Institution provider candidates

   Pittsburgh
   Moffitt
   UCLA
   Mayo
   NYU

## 5. Sample size justification:

Although the screening contexts for Reference Sets A and B have been specified, there are still many different conditions (sensitivities, specificities, and subpopulations) within those contexts in which a potential biomarker may be useful. To calculate sample sizes, we have identified conditions under which a potential marker has definite utility and determined the sample size to have the desired characteristics (power for a given type I error rate) in that condition. The choice of conditions and resulting test statistics used for the power analyses are not intended to place any restriction on the proposals for use of the specimens. Researchers applying for samples from the Reference Sets may anticipate different conditions and propose different test statistics than those assumed in these sample size calculations. However, as part of the application, we expect researchers to specify their analysis plan, state the conditions under which the marker will be considered to have successfully passed validation, and determine the power of their analysis given the numbers of samples in the requested Reference Set(s).

### 5.1 Rapid single biomarker pre-validation sets

*For Reference Set A*, the consensus of the clinicians in the LCPC was that a conservative estimate of the prevalence of disease in the screening population would be 1%. There was also consensus that a marker with a sensitivity of 0.80 and specificity of 0.70 would be worthwhile enough to merit further consideration, at least in the context of being included in a panel of markers evaluated in a combined pre-validation specimen set. To merit further consideration as a single marker, the Committee thought that a high sensitivity, say 0.85, was needed given the severity of the disease.

If considering all individuals at risk for lung cancer, the positive predictive value (PPV) would be 0.01, the same as the prevalence. For a biomarker to pass the rapid pre-validation phase, it will need to increase the PPV by an amount that is clinically significant. To compute the needed sample size, we specify two PPVs:

- $PPV_1$ is a PPV that, although better than 0.01, is not sufficiently better than 0.01 to be of clinical utility; we will choose a sample size that will accept such a PPV to move forward with low probability.
- $PPV_2$ is a PPV that is sufficiently better than 0.01 that it would have definite clinical utility; we will choose the sample size that will accept such a PPV to move forward with high probability.

For the computations here, we choose $PPV_1$ to be 0.013. $PPV_2$ is calculated using sensitivity of 0.80, specificity of 0.70, and disease prevalence of 0.01, and is equal to 0.026. Thus, in the screening population we conservatively assume that one in 100 individuals screened will have lung cancer. The consensus of the Committee members is that a test that increases that frequency to one in 40 screenees would be clinically worthwhile as a possible component in a marker panel for detection of lung cancer. A test that increases the frequency to one in 80 screenees would not be a large enough gain to be clinically worth further investigation as a candidate marker in a panel of markers. The specific combinations of sensitivity and specificity that we will design our sample size around are therefore:

| | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|
| Definite clinical utility | 0.80 | 0.70 | 0.026 | 0.997 |
| Insufficient sensitivity | 0.40 | 0.70 | 0.013 | 0.991 |
| Insufficient specificity | 0.80 | 0.40 | 0.013 | 0.995 |
| No screening test | -- | -- | 0.010 | 0.990 |

To compute the needed sample size of cases, we determine the sample size for the exact test of a binomial proportion with the property that the observed sensitivity either does not differ from the insufficient sensitivity level (0.40) at a p-value of $p_1$ or does not differ from the definite clinical utility sensitivity level (0.80) at a p-value of $p_2$, but not both. We choose the number of controls in a similar manner using the desired specificity. Intuitively, this means that either the marker result might be no better than a marker of little clinical utility (and hence it's not a good candidate to move forward) or the marker result might be as good as a marker of definite clinical utility (and hence should be moved forward to the next step in validation). The values for $p_1$ and $p_2$ are chosen to have the desired power for $PPV_1$ and $PPV_2$. We use $p_1 = 0.025$ and $p_2 = 0.025$. With these values, the needed sample sizes are 27 cases and 50 controls.

*For Reference Set B*, in five lung cancer screening trials(*3-7*), the prevalence of lung cancer among screened individuals with an abnormal CT result ranged from 2.8% to 11.5% , with only the Henschke study showing a rate of greater than 5%. To compute the positive predictive value (PPV) of markers in the screening setting, we will assume a prevalence of disease of 5%. The PPV of CT alone is thus 0.05. The Committee agreed that in this screening context, a sensitivity of 0.80 and specificity of 0.70 (with resulting PPV of 0.123) had sufficient clinical utility, at least for inclusion in a panel of markers. For a stand-alone marker, a sensitivity of at least 95% was thought necessary. A PPV of 0.075 was deemed not sufficiently better than CT alone to merit further evaluation. Thus, we compute sample size to distinguish among the following three scenarios:

| | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|
| Definite clinical utility | 0.80 | 0.70 | 0.123 | 0.985 |
| Insufficient sensitivity | 0.46 | 0.70 | 0.075 | 0.961 |
| Insufficient specificity | 0.80 | 0.48 | 0.075 | 0.978 |

Using the same method of determining sample size as discussed for Reference set A, we calculate that we need 38 cases and 87 controls in Reference Set B.

### 5.2 Panel of biomarkers pre-validation set

This set will be used to **construct** and **test** a panel of biomarkers for diagnosis of lung cancer. We presume that any individual marker alone will not have adequate performance to merit a multi-center validation study. To construct and test a panel of markers, EDRN and Lung Cancer SPORE collaborators will select a list of biomarker candidates. The assays for these candidates will be performed on the training set and the diagnostic rule will be established. That rule, if it meets the criterion for access (that means the completion of Phase I (*8*)), will then be validated on the validation set. Validation of individual markers using these sets may skip this first step if there is sufficient evidence, as agreed upon by the EDRN and Lung SPORE collaborators, that they meet the criteria for access. The confirmation on the validation set represents a successful completion of Phase II validation (if the assays are sufficiently developed to be robust and can be used in clinical setting). The sample size required here is much bigger than that for rapid pre-validation because of the requirement to demonstrate sufficient sensitivity and specificity to satisfy Phase II validation.

Sample size and power calculations are not applicable for the classifier **construction** phase (training sample) because the power depends on the signal to noise ratio, the nature of the data and analytical method. The sample size of 150 cases and 150 controls is based on the feasibility and cost, and the fact that this size is as large or larger than most of the biomarker studies reported.

### 5.3. Phase II validation set

*For Reference Set A*, the power calculation for the **validation** phase is based on the following assumptions:
1) Given the severity of lung cancer, we assume that a marker must have sensitivity of at least 0.80 and specificity of at least 0.80 to have clinical utility, so a marker with sensitivity or specificity below these thresholds would fail to pass this threshold. We choose our sample size so that a marker with sensitivity above 0.90 and specificity above 0.90 would have high probability of passing the validation step.

2) Power is calculated based on joint tests for sensitivity and specificity and adjusted for this multiplicity. We utilized two-sided tests of significance at a p-value of 0.05 and require approximately 90% power to distinguish the alternatives described in 1).

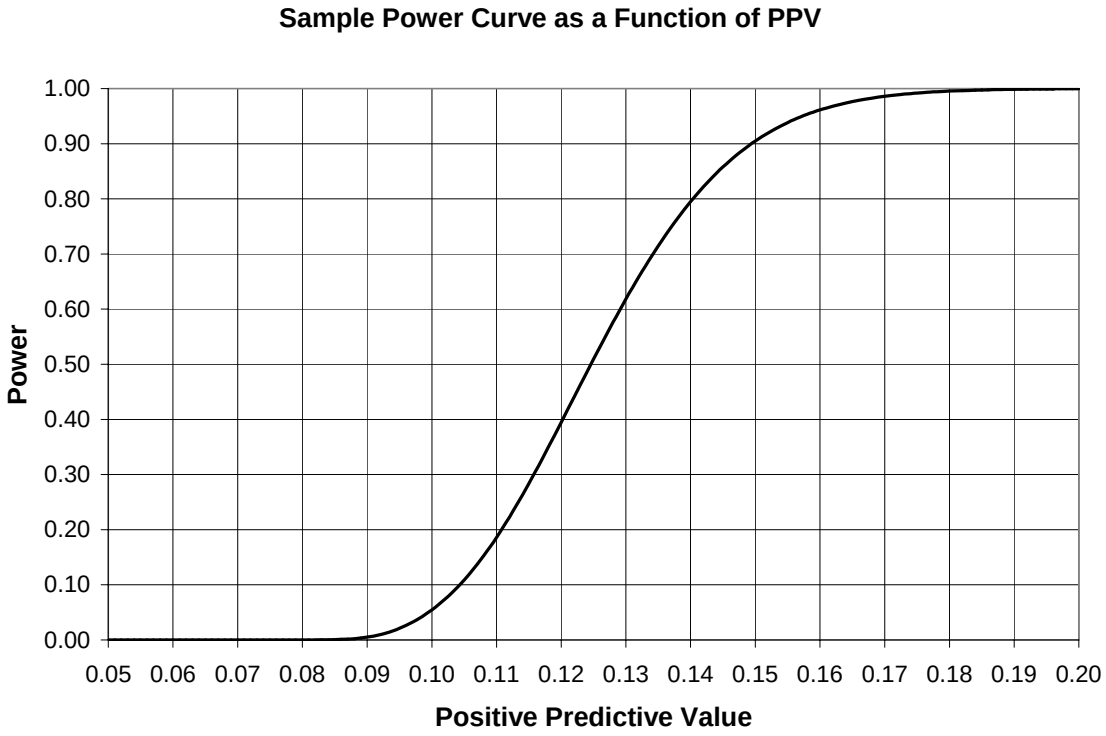
With these assumptions, the sample size needed is 180 cases and 180 controls. The positive and negative predictive values for the scenarios described above are shown below, under assumptions that the prevalence of lung cancer in the screened group is 0.01 (the conservative estimate used for the rapid pre-validation set).

| | | | Prevalence = 0.01 | |
|---|---|---|---|---|
| | Sensitivity | Specificity | PPV | NPV |
| Definite clinical utility | 0.90 | 0.90 | 0.083 | 0.9989 |
| Insufficient sensitivity | 0.80 | 0.90 | 0.075 | 0.9978 |
| Insufficient specificity | 0.90 | 0.80 | 0.043 | 0.9987 |

*For Reference Set B*, the power calculation for the **validation** phase is based on the following assumptions:
1) We assume that a marker which, for some cutpoint, has sensitivity of 50% and specificity of 85%, with resulting positive predictive value of 0.15, has clear clinical benefit, while a marker with sensitivity of 40% and specificity of 75%, and resulting PPV of 0.078, is clinically unacceptable. When calculating PPV and NPV, we assumed that in the screening population, the prevalence of lung cancer is 0.05.
2) Power is calculated based on joint tests for sensitivity and specificity and adjusted for this multiplicity. We utilized two-sided tests of significance at a p-value of 0.05 and require approximately 90% power to distinguish the alternatives described in 1).

With these assumptions, the sample size needed is 250 cases and 170 controls. Figure 4 demonstrates a sample power curve as a function of PPV for this sample size (it is only a sample since the power depends not just on PPV but also on the combination of sensitivity and specificity.

**Figure 4**




The entire combined pre-validation Reference Set (either A or B) may also be used in lieu of a validation study if a marker or panel of markers already has a decision rule developed on independent data and the investigator wishes to validate the marker on the standard specimen reference set. In this case, the marker would be evaluated on both the training and test sets together with the pre-chosen decision rule and the entire combined set being considered a test set. Thus, for Reference Set A, there would be 330 cases and 330 controls analyzed, while for Reference Set B there would be 320 cases and 400 controls analyzed. The table below shows the differences in sensitivity or specificity detectable using these sample sizes.

Sensitivity/specificity detectable with 90% power

| | "Bad sensitivity/specificity" | | | |
|---|---|---|---|---|
| Sample size | .800 | .850 | .900 | .950 |
| 320 | .876 | .917 | .954 | .986 |
| 330 | .875 | .916 | .953 | .985 |
| 400 | .869 | .910 | .949 | .983 |

## 6. Required and desired Common Data Elements (CDEs):

The EDRN has developed Common Data Elements (CDEs) for use in EDRN studies to enhance the ability of studies to share information through the use of identical data elements. It is recognized that retrospective specimens included in these reference serum/plasma sets from different institutions most likely will not have been collected using these EDRN CDEs. For this reason, not all of the established EDRN Core CDEs will be required for the specimens banked in this protocol. **Appendix E** indicates the required and desired CDEs to be collected to annotate the individual from whom a specimen was obtained and indicates the CDEs (all required) needed to annotate the specimen itself. In the event the specific EDRN CDEs were not collected, the EDRN Data Management and Coordinating Center (DMCC) will work with the investigators at each contributing institution to map their data elements to the respective EDRN CDEs. Once this mapping is completed, the site providing specimens is then responsible for providing the requisite data, stored using the EDRN CDE variable names and formats.

## 7. Sample collection and processing requirements:

- Approximately equal numbers of cases and controls will be included from each of the contributing institutions when possible
- All samples collected contemporaneously at all sites within the past 5 years (since 2000)
- Sample collection: for samples already collected, a copy of the collection protocol is requested from each participating institution. For prospectively collected cases and controls please refer to the suggested standard operating procedure (**Appendix D**)
- Shipping of minimum of 1-2 mL of serum and plasma
- Shipping of sample after FIRST freeze (meaning sample was frozen once at each institution and shipped frozen. Samples will be thawed and aliquoted at the Frederick site.
- Aliquots: 8 x 100 μL
  8 x 25 μL.

It is a good idea to group any samples that go to different sets and label them according to which set they belong (Set A, serum / Set A, plasma / Set B, serum / Set B, plasma). Shipping samples on dry ice OVN delivery to NCI Frederick Facility and have their FEDEX account number charged for shipping (2649-0814-0). Address the package to:

Rich Buxton
Fisher BioServices
4600 Wedgewood Blvd., Suite H
Frederick, MD 21703

It is best to contact Rich Buxton before sending samples. He can be reached at 301-694-5911 and his email address is rich.buxton@fishersci.com. He needs to be notified before samples arrive so that he is prepared to receive and process them. On the shipping package in the Customer Reference field indicate that these samples are for the "EDRN/SPORE Lung Sets". Inform Rich Buxton of the tracking number after the package has been sent out.

Recommended packages for shipping can be purchased at http://www.saftpak.com/
If undamaged, these containers can be shipped back to the contributing institution. Sufficient dry ice should be enclosed to last for 2 days. Shipping should be done on a Monday or Tuesday to guarantee delivery to the Frederick facility before Friday of the same week.

## 8. Storage conditions:

*Consensus on a repository at NCI-Frederick:*
The Committee agreed on initially developing two serum/plasma reference sets (Reference Sets A and B]. Tissue samples and other biological specimens will be addressed in the future after showing feasibility with blood samples.

*Centralization of the 1-2 mL serum/plasma samples at the NCI Frederick facility:*
For oversight and monitoring Fisher BioServices at the NCI Frederick Facility provides as part of the monthly maintenance cost for the freezer the following services: They ensure contents of the freezer will be kept at $-80^{o}$C and have sufficient backup freezers and power backup to guarantee this. Karl Krueger (NCI/CBRG) will be the chief NCI/Committee contact for the repository samples, but in his absence Peter Ujhazy (NCI/SPORE) and Sudhir Srivastava (NCI/CBRG) can also contact Fisher BioSciences for shipping of Reference Set samples, i.e. who can have access to which samples given approval by the LCPC designated Specimen Sharing Committee.

## 9. Process for evaluation of biomarkers and distribution of samples:

A The ***Specimen Sharing Committee*** was created within the LCPC and has drafted a process through which the Reference Sample Sets could be accessed (**Appendix C**). This Committee is charged with review of requests for either or both of the Reference Sample Sets. Common guidelines and procedures were developed. This Committee includes: Drs. Wiest (Chair), Ujhazy, Krueger, Bigbee, Massion, Franklin, Rom and Thornquist.

*1. Review criteria:*
Review criteria are based on scientific merit and compatibility with LCPC objectives. Six formal criteria are used to assess the suitability of proposals for access to aliquots of the Reference Sample set(s):

- Scientific merit
- Study design: relevance of the retrospective Reference Sets A and B
- Technical parameters: reproducibility, sensitivity, specificity, throughput, automation
- Clinical or scientific impact: e.g., more common cancers or a significant impact in less common neoplasia
- Practicality and feasibility: e.g., cost, required sample size, amount of biospecimen required
- Collaborative strength, including contribution of resources and technology.

*2. Review process:*
The LCPC Specimen Sharing Committee will review all applications for access to the Reference Sample Sets.
The review process is described below:
1. Copies of proposals received by the receipt date are forwarded from the LCPC Program Office to the members of the Specimen Sharing Committee within a week after the application receipt date (web-based electronic review system).
2. The LCPC Specimen Sharing Committee evaluates and scores applications and sends results of the review to the LCPC. The evaluation is expected to be complete within one month following the application receipt date.

3. The LCPC renders final approval by majority vote of the reviewed proposals and communicates decisions to the NCI Frederick Facility for release of the appropriate samples. These actions are expected to occur within three months after the application receipt date.

*3. Additional information for validation study proposals:*
Progress of a biomarker to a full validation study is a critical step in the development of a biomarker and is, therefore, a critical priority of the LCPC.
Proposal: See **Appendix B**
A pre-proposal/letter-of-intent, limited to 3 pages, must be submitted to the LCPC Program Office. Validation studies are collaborative, therefore, the pre-proposal must name the EDRN sites which will be included as part of the collaboration as well as the EDRN DMCC. An EDRN Clinical Epidemiology and Validation Center (CEVC) should be included as needed. The full proposal should not exceed 10 pages. A detailed background and rationale are not necessary but presentation of preliminary data documenting the performance of the proposed marker is required.
Full proposals are reviewed monthly by the Specimen Sharing Committee. Submissions received by the LCPC Program Office by the first of the month are reviewed that month. by the EDRN Executive Committee (EC) .

Results of the pre-validation study will be made available to the LCPC for review and comparative/combined analysis with data from other biomarker pre-validation studies using the same Reference Set samples. This process includes a requirement for submission by the investigators of their primary data for confirmatory analysis by the EDRN DMCC. Details of these guidelines are provided in **Appendix C**.

## 10. IRB/Material Transfer Agreement (MTA):

*For investigators providing samples to the Reference Sample Sets in the repository:*
Copy of the consent form and copy of the protocol with approval date.

- Samples collected with IRB approved consent form.
- Consent informs individuals that the samples will be shared with other investigators and to investigate eventually other types of cancer and other disease
- None of the 18 HIPAA identifiers will be shared, all information is de-identified.

*For investigators requesting access to Reference Sample Sets in the repository*
IRB approval for use of reference set samples from the repository

*Material Transfer Agreement(MTA):*

Proposed format: Standard format provided in the **Appendix F**.

**11. Summary:**

| | **Reference Set A** | **Reference Set B** |
|---|---|---|
| Clinical question | Diagnosis of lung cancer | Early diagnosis of lung cancer |
| Study design | Case-control study | Case-control study |
| Study population | All suspicious lung lesions on CXR or on CT | CT screening |
| Cases | Lung cancers, ≥50% Stage I | Detected by CT- Lung cancer ≥0.5cm and ≤3cm |
| Controls | High risk individuals as defined by ≥50 YO, ≥ 30 PKYs of smoking<br>All lung lesions on CXR or on CT suspicious for lung cancer but proven not to be cancer at 1 year follow up<br>75 patients with other cancers (25 breast, 25 colon, 25 prostate) | Detected by CT with a lung nodule ≥0.5cm and ≤3cm<br>CT screened without a lung nodule (10% of total # of controls)<br>All free of disease at the 1 year F/U CT75 patients with other cancers (25 breast, 25 colon, 25 prostate) |
| Matching criteria | Age, sex, smoking status, PKYs | Age, sex, smoking status, PKYs |
| Sample size<br>(1) Single biomarker pre-validation<br><br>(2) Panel of biomarkers pre-validation<br><br>(3) Phase II validation | <br>(1) 27 cases 50 controls ± 25 other cancer controls<br><br>(2) 150 cases and 150 controls<br><br>(3) 180 cases and 180 controls + 75 other cancer controls | <br>(1) 38 cases 87 controls ± 25 other cancer controls<br><br>(2) 150 cases and 150 controls<br><br>(3) 250 cases, 170 controls + 75 other cancer controls |
| Pulmonary diseases | Enriched across whole population | Enriched in controls with CT detected lung nodules |
| Institution provider candidates | Pittsburgh<br>Vanderbilt<br>MDACC<br>Moffitt<br>NYU<br>UCLA<br>UCHSC<br>JHU | Pittsburgh<br>Moffitt<br>UCLA<br>Mayo<br>NYU |

## 12. Appendix:

# A. Repository description

**Reference Set A**

| update 6/20/2005 | UCHSC | | Pittsburgh | | Vanderbilt | | Moffitt | | UCLA | | NYU | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Blood samples | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| Serum | | X | X | | X | | | X | | X | X | | |
| Plasma | X | | X | | X | | X | | X | | X | | |
| 1-2 mL available | X | | X | | 1 | | 1 | | X | | 1 | | |
| Serum collection | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| Uniform protocol cases/controls | | X | X | | X | | | X | | X | X | | |
| Frozen within 4 hours | | | X | | X | | | X | | X | X | | |
| First thaw available | | | X | | X | | | X | | X | X | | |
| # cases available | | | 50 | | 50 | | | X | | X | 50 | | **150** |
| # matched controls available | | | 50 | | 50 | | | X | | X | 50 | | **150** |
| # breast cancer controls | | | 10 | | | X | | X | | X | | X | **10** |
| # colon cancer controls | | | 10 | | | X | | X | | X | | X | **10** |
| # prostate cancer controls | | | 10 | | | X | | X | | X | | X | **10** |
| Plasma collection | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| Uniform protocol cases/controls | X | | X | | X | | X | | X | | X | | |
| Frozen within 4 hours | X | | X | | X | | NOT RECORDED | | X | | X | | |
| Two spins (platelet contamination) | X | | X | | | X | | X | | X | | X | |
| First thaw available | | X | X | | X | | X | | X | | X | | |
| # cases available | 50 | | 50 | | 50 | | 10 | | 50 | | 50 | | **260** |
| # matched controls available | 50 | | 50 | | 50 | | 10 | | 50 | | 50 | | **260** |
| # breast cancer controls | | | 10 | | | X | | X | | X | | X | **10** |
| # colon cancer controls | | | 10 | | | X | | X | | X | | X | **10** |
| # prostate cancer controls | | | 10 | | | X | | X | | X | | X | **10** |
| Clinical Data Elements available | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| | X | | X | | X | | X | | X | | X | | |
| LCPC request conforms to IRB consent | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| | X | | X | | X | | X | | X | | X | | |

**Reference Set B**

| update 6/20/2005 | UCHSC | | Pittsburgh | | Vanderbilt | | Moffitt | | UCLA | | NYU | | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Blood samples | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| Serum | X | | X | | | X | | X | | X | X | | |
| Plasma | X | | X | | | | X | | | | X | | |
| 1-2 mL available | X | | X | | | | X | | | | X | | |
| Serum collection | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| Uniform protocol cases/controls | X | | X | | | | | X | | X | X | | |
| Frozen within 4 hours | X | | X | | | | | X | | | X | | |
| First thaw available | X | | X | | | | | X | | | X | | |
| # cases available | 50 | | 40 | | | | | | | | 10 | | **100** |
| # matched controls available | 50 | | 100 | | | | | | | | 50 | | **200** |
| # breast cancer controls | | | 10 | | | | | | | | | | **10** |
| # colon cancer controls | | | 10 | | | | | | | | | | **10** |
| # prostate cancer controls | | | 10 | | | | | | | | | | **10** |
| Plasma collection | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| Uniform protocol cases/controls | X | | X | | | | X | | | X | X | | |
| Frozen within 4 hours | X | | X | | | | NOT RECORDED | | | | X | | |
| Two spins (platelet contamination) | X | | X | | | | | X | | | | X | |
| First thaw available | X | | X | | | | X | | | | X | | |
| # cases available | 50 | | 40 | | | | 50 | | | | | | **140** |
| # matched controls available | 50 | | 100 | | | | 50 | | | | | | **200** |
| # breast cancer controls | | | 10 | | | | | X | | | | | **10** |
| # colon cancer controls | | | 10 | | | | | X | | | | | **10** |
| # prostate cancer controls | | | 10 | | | | | X | | | | | **10** |
| Clinical Data Elements available | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| | X | | X | | | | X | | | | X | | |
| LCPC request conforms to IRB consent | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | Yes | No | |
| | X | | X | | | | X | | | | X | | |

## Appendix B. Study application form and scientific proposal

| *Early Detection Research Network*<br>*NCI/EDRN/SPORE Lung Cancer Proteomics Committee*<br>Part 1: Request For Reference Sample Sets | |
|---|---|
| Date of Submission: | |
| Investigator:<br><br>Name:<br>Institution:<br>Address: | <br><br>Phone:<br>Fax:<br>E-mail |
| Specimen Reference Set(s) Requested | |
| Reference Set Requested<br><br>*Rapid Pre-validation Set*<br>☐ *Rapid Pre-validation Set A*<br>☐ *Rapid Pre-validation Set B*<br>*Combined Pre-validation Set*<br>☐ *Combined Pre-validation Set A*<br>☐ *Combined Pre-validation Set B* | Specimen Type<br><br>☐ Serum<br>☐ Plasma |
| | |
| Minimum volume of each sample required:<br>(microliters) | Expected length of study:<br>months |
| Institutional Approval | |
| Do you have IRB approval to work with the requested samples? | ☐ Yes: Institution:<br>Approval number:<br>☐ No<br>☐ Pending: Expected date: |
| Funding | |
| How will testing of the Reference Sample Set(s) be funded? | ☐ Current NIH-funded grant:<br>Grant No.<br>Annual Direct Costs:<br>Funding Period:<br><br>☐ Other sponsorship: Please provide a letter of commitment from the sponsoring agency, company, or foundation.<br><br>☐ Other: specify: |

# Part II: Scientific Proposal

Using the standard PHS 398 Continuation Page (http://grants.nih.gov/grants/funding/phs398/continuation.doc) address the following items as outlined below (3-5 pages recommended)

I. **Clinical Relationship**: Clearly state the clinical question that you are trying to address: early detection or diagnosis. If other clinical questions are to be addressed (e.g., risk assessment or prognosis) provide a rationale. How would the Reference Set(s) samples expedite addressing the intended clinical question?

II. **Background and Significance**: Clearly state the scientific rationale of the proposal for using the requested Reference Set(s) samples. Describe your biomarker/platform and how you came upon its discovery/development for potential application in cancer detection.

III. **Preliminary Data & Methods:** Provide sufficient information describing how experiments were performed, details on convenience discovery samples used, and presentation of data in terms of specificity, sensitivity, and variance of your measurements. Explicit description of your studies will facilitate review considerations. Figures and other supporting documentation can be appended to your proposal. The application is expected to contain at least preliminary analysis of lung cancer samples.

IV. **Data Analysis Plan:** Provide adequate detail concerning how statistical analysis of your data generated from the Reference Set(s) samples will be performed and a justification that the requested References Set(s) is/are large enough to demonstrate the utility of the biomarker. Describe the statistical resources at your disposal. If you require statistical support, the EDRN/SPOREs can assist you with this.

V. **Collaboration:** In this section state your willingness to deposit all primary data obtained using the Reference Set(s) samples with the EDRN Data Management and Coordinating Center (DMCC). EDRN/SPORE programs may compare this data as a reference with other biomarkers applied to the same samples.

VI. **Future Plans**: If the biomarker is found to have promising performance characteristics, the EDRN/SPORE programs might be interested in working with you to proceed to Phase II clinical validations. Address each specific scenario below according to your intentions:

   a. Do you plan to approach EDRN/SPOREs for funding and collaboration in proceeding to a Phase II validation study? If not, do you have other resources where validation can be accomplished? Describe clearly other resources at your disposal and how they are sufficient to complete a larger Phase II validation study if you will not seek funding support from the NCI.
   b. Are you amenable to working within the collaborative framework of EDRN/SPOREs in proceeding to Phase II studies?
   c. If deemed beneficial, will you be amenable to including your biomarker into a larger panel of biomarkers for Phase II validation conducted by other NCI programs?
   d. If refinements will improve the performance of the biomarker test, will you concur with further development of the test? Will it be advantageous to include resources of the EDRN for this purpose?

# Appendix C. Reference Sample Set Sharing Guidelines

## NCI/EDRN/SPORE Lung Cancer Proteomics Committee Reference Sample Set Sharing Guidelines

The goal of the NCI/EDRN/SPORE Lung Cancer Proteomics Committee (LCPC) is to develop the requisite sample resources to pre-validate serum biomarkers for the early diagnosis of lung cancer. These sample sets will be maintained at the NCI-Frederick repository. Researchers may use these resources and process for testing the reproducibility of biomarkers across institutions and analytical platforms. The Specimen Sharing Committee was created within the LCPC and has designed an application and review process through which these Reference Set samples could be accessed. This Committee is charged with review of requests for either or both of the Reference Set samples. Common guidelines and procedures are being developed and implemented. Applications for multiple Reference Set samples will be considered based on the scientific merit of the application.

**The patient population context:**

*1. Reference Set A* will focus on pre-validation of biomarkers of *diagnosis of lung cancer* and target lung cancer diagnosed for individuals at high risk for lung cancer or abnormal chest x-ray CXR or chest CT but outside of the context of a CT screening trial. **The clinical question** to be tested after pre-validation relates to whether a serum/plasma biomarker has added value to current clinical tests (CT scan and/or PET scan) for the diagnostic evaluation of pulmonary nodules and to whether such a biomarker could reduce the number, and the attendant cost, of unnecessary invasive tests (PET or tissue biopsy) or futile thoracotomies.

*2. Reference Set B* will focus on pre-validation of biomarkers of *early diagnosis (screening) of lung cancer* and targeting a specific population of lung cancer patients diagnosed in the context of a computed tomography (CT)-based screening trial of high risk individuals. **The clinical question** to be tested after pre-validation relates to whether a serum/plasma biomarker has added diagnostic value to current tests (CT scan and/or PET scan) for the diagnostic evaluation of CT-detected pulmonary nodules.

*3. Controls*: The controls include blood samples from groups of individuals with matched age, sex, race, smoking status and smoking pack year history of smoking to the lung cancer cases. In addition, these controls include samples from subjects with non-malignant lung disease and malignant disease from other organ sites.

**The biomarker context:**

**1.** *Rapid single biomarker Pre-validation Reference Set:* A biomarker with demonstrated good performance in a discovery sample set, typically from one institution, needs to be tested in samples from *independent clinical populations* before going forward to validation in the combined pre-validation set or in a multi-center validation study.

**2.** *Panel of biomarkers Pre-validation Reference Set:* The performance of an individual biomarker is insufficient alone but may have utility when *combined in a panel* with other biomarkers measured in the same sample set. In this evaluation, investigators will establish (in the "construction set") a combination marker decision rules. In the second set of specimens (the "test" set) they will evaluate the constructed rule in an independent sample of specimens to determine the rule's operating characteristics (e.g., sensitivity, specificity, positive and negative predictive value). Since all researchers utilizing the

combined pre-validation set will be evaluating their markers on essentially identical sets of specimens and contributing the data from their assays to the central database of assay results, over time the reference set will become progressively more valuable as the number of possible marker combinations that may be constructed and tested using this reference set expand.

**3.** *Phase II biomarker Validation Reference Set:* For single or combined biomarkers who have passed **1.** or **2.**

**Sample collection and standardization:** A critical aspect of serum biomarker validation is to be able to confirm the predictive value of biomarkers beyond the heterogeneity in sample collection, preparation, storage and shipping. The focus initially is on obtaining retrospective samples from institutions where cases and controls were collected, processed, and stored under the same protocol. **See Appendix D**.

**Review process:**

All applications will be reviewed by the Specimen Sharing Committee and the review will be based on the scientific merit of the application. After providing specific details related to the sample set(s) being requested and institutional approval to use these sets, the investigator requesting access is then expected to address the following topics as provided on the application form in relation to his/her biomarker and future intentions. The application is expected to contain at least preliminary analysis of lung cancer samples.

Clinical Relationship

Background and Significance

Preliminary Data & Methods

Data Analysis Plan

Collaboration

Future Plans

In essence, the Specimen Sharing Committee will review the applications prior to granting access to Reference Set samples. These criteria are established by the LCPC before the Reference Set samples become available. For each review conducted, it is expected that an adequate biostatistical critique will be provided by involvement of the EDRN Data Management and Coordinating Center (DMCC) or SPORE statistical group to ensure that appropriate consideration is given to statistical concerns of the proposal.

Upon receiving an inquiry or request regarding access to the Reference Set(s) samples the NCI Program staff committee member will be notified to send an application form and any other relevant documents to the investigator. After the completed application has been returned, the NCI Program staff committee member will then forward it to the members of the Specimen Sharing Committee. The Committee, in a timely manner (within one month), will review and discuss the application and offer a recommendation of whether 1) the investigator should be sent the requested Reference Set(s) samples, 2) further clarification or revision are needed, or 3) the request is considered low priority and deferred.

1) If approval is given, the LCPC will be notified at its next meeting (or by email if extenuating circumstances arise) by the Specimen Sharing Committee Chair (or Co-chair). In principle, the LCPC will concur with all approvals recommended by Specimen Sharing Committee unless special issues are raised. NCI Program Staff will then notify the repository facility in Frederick to prepare the materials needed for sending the appropriate Reference Set(s) samples.

2) If the application is unfunded at the time the request is submitted and is approved, the LCPC will provide to the investigator a letter of commitment. This obligation of samples will last for one year from the approval date.

3) If further clarification is needed, the LCPC will inform the NCI Program staff committee member what concerns or questions remain with the application. The NCI Program staff committee member will then communicate with the investigator of these issues to ask for a resubmission.

4) If the request is deemed low priority and deferred, the LCPC will provide the rationale to the NCI Program staff committee member why the request was deferred. This staff member will then relay this decision and its reasons to the investigator requesting access.

**Timeline:**

It is expected that the review of applications will take approximately one month following submission of a complete application. If the application is approved, the requested Reference Set(s) samples will be sent to the applicant within two weeks of the approval date.

**Contacts:**

Karl Krueger, Ph.D.
Program Director
Cancer Biomarkers Research Group
Division of Cancer Prevention
National Cancer Institute
6130 Executive Plaza North, Suite 3136
Rockville, MD 20852
Telephone: (301) 594-1044
Fax: (301) 402-8990
Email: kruegerk@mail.nih.gov

Peter Ujhazy, M.D., Ph.D.
Program Director
Organ Systems Branch
OCTR, ODDES, NCI
6116 Executive Blvd. Room 7015
Rockville, MD 20852
Telephone: (301) 496-8528
Fax: (301) 402-5319
Email: pu5s@nih.gov
http://spores.nci.nih.gov

# Appendix D. Specimen collection recommendations

## NCI/EDRN/SPORE Lung Cancer Proteomics Committee

Adapted from NCI/FDA Proteomics Program Protocol

**Standard Operating Procedure**

*Collection of Serum and Plasma Samples for Proteomic Analysis*

### I. Principle

The collection of human blood samples for proteomic analysis requires that patient sample collection, storage, transport and handling remain consistent within rigid guidelines for optimal results.

### II. Materials

1. Requisition
2. Gloves
3. Sharps Container
4. Vacutainer needles, 20-22g and vacutainer hub
   or Butterfly needle, attached tubing and Luer adapter
5. Tourniquet
6. Antiseptic wipes
7. Red top glass, no additive, no clot activator with uncoated interior, vacutainer tubes (examples: BD366397, BD366442, BD366441).
8. Bandages
9. Centrifuge
10. Polypropylene screw top freezer tubes, 2 ml.

### III. Specimen

The specimen of choice for proteomic analysis at the NCI/FDA Proteomics Program is serum obtained from whole blood collected in red top vacutainer tubes with no additives or clot activators. The blood specimen should be allowed to clot for 40 - 50 minutes. The serum should be transferred to a transfer/storage tube **within 2 hours of collection**. Samples are then frozen at $-70^0$C until testing is performed. The serum should be free of hemolysis and clots. Freezing and thawing cycles should be kept to a minimum. (Plasma samples can also be analyzed however every sample in the set must be plasma and the same anticoagulant used for all samples).

### IV. Procedure

1. Identification of the patient is crucial. The person obtaining the blood specimen must ensure that the blood specimen being drawn is from the individual designated on the request form

2. Assemble the supplies to be used in obtaining the specimen. Label the tubes.

3. Put on disposable gloves.

4. The patient should be comfortably seated in a venipuncture chair. The arm should be positioned on a slanting armrest in a straight line from the shoulder to the wrist. The arm should not be bent at the elbow.

5. Apply a tourniquet 2 inches above the antecubical fossa or above area to be drawn with enough pressure to provide adequate vein visibility. Have the patient form a fist. Select the site for venipuncture.

6. Clean the forearm of the patient with antiseptic wipe in a circular motion beginning at the insertion site. Allow the antiseptic to dry.

7. Anchor the vein by placing the thumb 2 inches below the site and pulling the skin taut to prevent the vein from moving. The holding finger is placed below the site, not above, to prevent accidentally sticking the finger with the needle.

8. Using the dominant hand, insert either the vacutainer needle or the butterfly needle (if using vacutainer needle, attach hub first). Push the evacuated tube onto the vacutainer hub or the Luer adapter if using a butterfly.

9. Release the tourniquet once blood flow is established.

10. Carefully remove the tubes when full without dislodging the needle. The tube will automatically stop filling when the vacuum is gone leaving the tube approximately three-fourths full.

11. Lightly place a sterile gauze pad over the venipuncture site. Gently remove the needle.

12. Apply pressure to the site with sterile gauze. Apply bandage. Instruct the patient to leave the bandage on for at least 15 minutes.

13. Dispose of the needle in a sharps container.

14. Remove gloves and wash hands.

**Plasma and buffy coat:**

1. Under the direction of a qualified and licensed physician, trained phlebotomists will collect blood from each donor into vacutainer that contain either no anticoagulant or an appropriate volume of anticoagulant $K_2$EDTA, to prevent clotting.

2. From each consenting donor, **30 mL** of whole blood will be collected.

3. The specimens are centrifuged immediately after blood draw at 1000 g (RCF) for 10 minutes at 4° C. The resultant plasma (assume 40% yield) is transferred into secondary centrifuge tubes.

4. Save the Buffy coat at -20°C for DNA extraction. No need to transfer buffy coat in another tube. DNA can be extracted from WBC after RBC lysis.

5. The secondary tubes are then centrifuged at 1500 g (RCF) at 4º C for 5 minutes to remove all potentially remaining cells.

6. Aliquots will be transferred into labeled cryovials and frozen at below -80° C within 2 hours of processing.

**Serum:**

1. Approximately 5ml of blood is collected in a sterile vacutainer (recommended BD Bioscience 366431, red top no additive vacutainer)

2. Leave on the bench at room temperature for ~ 45 minutes to allow clot to form.

3. Spin at 1000g for 10 min in a refrigerated ($4^oC$) centrifuge.

4. Transfer supernatant carefully into a polypropylene Eppendorf-style microfuge tube and store immediately at $-80^o$ C.

Dispose of all tubes and materials used to transfer patient samples in biohazardous waste.

## V. Procedural notes:

1. Do not draw from an IV, mastectomy or shunt arm.

2. Samples may also be drawn into a syringe and then dispensed into a vacutainer tube (before clotting) for processing.
3. ORDER OF DRAW: Blood collection tubes must be drawn in a specific order to avoid cross-contamination of additives between tubes. The recommended order of draw is:
   First - non-additive tube (red stopper or SST)
   Second - EDTA (lavender stopper)

   NOTE: Tubes with additives must be thoroughly mixed. Erroneous test results may be obtained when the blood is not thoroughly mixed with the additive.

# Appendix E. Clinical and Sample data elements

## E.1. REQUIRED

***Group Name: Lung Reference Set***
***CDEs required for one or more study groups***

---

795 Proposed study group
***Required***

- ☐ 1 Set A Cases: Lung cancer (CXR)
- ☐ 2 Set A Cases: Lung cancer (CT)
- ☐ 3 Set A Controls: High Risk
- ☐ 4 Set A Controls: Suspicious lung lesions
- ☐ 5 Set A Controls: Other cancers
- ☐ 6 Set B Cases: Lung cancer (CT screening)
- ☐ 7 Set B Controls: High risk
- ☐ 8 Set B Controls: CT nodule
- ☐ 9 Set B Controls: No nodule
- ☐ 10 Set B Controls: Other cancers

422 EDRN Site ID ⌊__⌋
***No mapping necessary, auto assigned by DMCC***

1063 Site Participant ID
***Required***

421 EDRN Participant ID ⌊_____⌋
***No mapping necessary, auto assigned by DMCC***

796 Date participant signed consent form ⌊__⌋ ⌊__⌋ ⌊____⌋ (Month, Day, Year)
***Required***

434 Gender (What is your gender?)
***Required***

- ☐ 1 Male
- ☐ 2 Female
- ☐ 9 Unknown/refused

436 Race (What is your race? Check all that apply.)
***Required***

- ☐ 1 White
- ☐ 2 Black or African-American
- ☐ 3 American Indian or Alaska Native
- ☐ 4 Asian
- ☐ 7 Native Hawaiian or other Pacific Islander
- ☐ 97 Other, specify:
- ☐ 99 Unknown/refused

793 Race (Other,specify)
***Required***

441 Ever smoke cigarettes regularly, at least one a day for a year or more? (Did you ever smoke cigarettes regularly, at least one a day for a year or more?)
***Required***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

443 Currently smoke at least one cigarette a day? (Do you currently smoke cigarettes regularly, at least one a day?)
***Required***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

1213 Average number of packs smoked per day? ⌊_____⌋
***Required***

444 Age quit smoking cigarettes? (How old were you when you permanently quit smoking cigarettes?)
⌊___⌋ Age
***Required***

1272 Living status:
***Required***

☐ 1 Alive with disease ☐ 2 Alive with no evidence of disease (NED)
☐ 3 Dead

1218 Last date known alive: ⌊__⌋ Month ⌊__⌋ Day ⌊____⌋ Year
***Required***

1048 Date of death ⌊__⌋ Month ⌊__⌋ Day ⌊____⌋ Year
***Required***

792 Ever had cancer [other than basal/squamous cell skin cancer] confirmed by a doctor? (Have you ever had cancer [other than basal/squamous cell skin cancer] confirmed by a doctor?)
***Required***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

451 Histologic confirmation?
***Required***

☐ 0 No ☐ 1 Yes

447 Cancer type/location
***Required***

☐ 1 Bladder ☐ 2 Bone
☐ 3 Brain ☐ 4 Breast
☐ 5 Cervix ☐ 6 Colon

☐ 7 Esophagus
☐ 8 Head & neck (mouth, nose, and throat)
☐ 9 Kidney
☐ 10 Liver
☐ 11 Leukemia
☐ 12 Lung
☐ 13 Lymphoma, including Hodgkins
☐ 14 Ovary
☐ 15 Pancreas
☐ 16 Prostate
☐ 17 Rectum
☐ 18 Skin (melanoma, no basal or squamous)
☐ 19 Stomach
☐ 20 Thyroid
☐ 21 Uterus
☐ 24 Testis
☐ 26 Vagina
☐ 97 Other, specify:
☐ 99 Unknown/refused
☐ 999 Unknown/refused

727 Cancer type/location (Other, specify)
***Required***

510 Date of diagnosis (MM/YYYY) Month Year
***Required***

574 Age at diagnosis Age
***Required***

1216 Date (MM/DD/YYYY) of diagnosis of lung cancer (histological or cytopathological report)
Month Day Year
***Required***

1217 Date of thoracotomy (MM/DD/YYYY):
***Required for cases, sets A and B***

1234 Chest X-ray date: Month Day Year
***Desired for Set A***

1235 Chest X-ray nodule size (cm):
***Desired for Set A***

1236 Chest X-ray nodule location:
***Desired for Set A***

1237 Chest CT date: Month Day Year
***Desired for Set A***

1238 Chest CT nodule size (cm):
***Desired for Set A***

1239 Chest CT nodule location:
***Desired for Set A***

1240 Reported adenopathy > 1 cm in the mediastinum at time of diagnosis?
***Desired for Set A***

786 Lung T-Stage, Pathologic
***Required****
***If Pathologic stage is not available, provide Clinical stage.***

☐ 2 pT0
☐ 3 pTis

☐ 4 pT1
☐ 5 pT2
☐ 6 pT3
☐ 7 pT4
☐ 8 pTX

789 Lung N-Stage, Pathologic
***Required****
****If Pathologic stage is not available, provide Clinical stage.***

☐ 1 pNX
☐ 2 pN0
☐ 3 pN1
☐ 4 pN2
☐ 5 pN3

790 Lung M-Stage, Pathologic
***Required****
****If Pathologic stage is not available, provide Clinical stage.***

☐ 1 pMX
☐ 2 pM0
☐ 3 pM1

785 Lung T-Stage, Clinical
***Required****
****If Pathologic stage is not available, provide Clinical stage.***

☐ 2 T0
☐ 3 Tis
☐ 4 T1
☐ 5 T2
☐ 6 T3
☐ 7 T4
☐ 8 TX

787 Lung N-Stage, Clinical
***Required****
****If Pathologic stage is not available, provide Clinical stage.***

☐ 1 NX
☐ 2 N0
☐ 3 N1
☐ 4 N2
☐ 5 N3

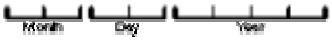
788 Lung M-Stage, Clinical
***Required****
****If Pathologic stage is not available, provide Clinical stage.***

☐ 1 MX
☐ 2 M0
☐ 3 M1

1257 Histologic type:
***Required***

☐ Squamous cell carcinoma
☐ Papillary squamous cell carcinoma

- ☐ 1 Clear cell squamous cell carcinoma
- ☐ 2 Small cell squamous cell carcinoma
- ☐ 3 Basaloid squamous cell carcinoma
- ☐ 4 Small cell carcinoma
- ☐ 5 Combined small cell carcinoma
- ☐ 6 Adenocarcinoma
- ☐ 7 Acinar adenocarcinoma
- ☐ 8 Papillary adenocarcinoma
- ☐ 9 Bronchioloalveolar carcinoma
- ☐ 10 Non-mucinous
- ☐ 11 Mucinous
- ☐ 12 Mixed mucinous and non-mucinous or indeterminate
- ☐ 13 Solid adenocarcinoma with mucin formation
- ☐ 14 Adenocarcinoma with mixed subtypes
- ☐ 15 Large cell carcinoma
- ☐ 16 Other
- ☐ 17 Unknown/refused
- ☐ 95

504 Date of follow-up data collection 


905 New primary cancer [other than basal/squamous cell skin cancer] confirmed by a doctor since last routine study contact? (Have you been diagnosed with a new primary cancer [other than basal/squamous cell skin cancer] since your last routine study contact?)
***Required***

- ☐ 0 No
- ☐ 1 Yes
- ☐ 9 Unknown/refused

742 Cancer type/location
***Required***

- ☐ 1 Bladder
- ☐ 2 Bone
- ☐ 3 Brain
- ☐ 4 Breast
- ☐ 5 Cervix
- ☐ 6 Colon
- ☐ 7 Esophagus
- ☐ 8 Head & neck (mouth, nose, and throat)
- ☐ 9 Kidney
- ☐ 10 Liver
- ☐ 11 Leukemia
- ☐ 12 Lung
- ☐ 13 Lymphoma, including Hodgkins
- ☐ 14 Ovary
- ☐ 15 Pancreas
- ☐ 16 Prostate
- ☐ 17 Rectum
- ☐ 18 Skin (melanoma, no basal or squamous)
- ☐ 19 Stomach
- ☐ 20 Thyroid
- ☐ 21 Uterus
- ☐ 24 Testis
- ☐ 26 Vagina
- ☐ 44 None
- ☐ Other, specify:
- ☐ Unknown/refused

97 999

728 Cancer type/location (Other, specify)
***Desired***

513 Histologic confirmation?
***Required***

☐ 0 No ☐ 1 Yes

578 Age at specimen collection ___ Age
***Required***

1064 Site Specimen ID
***Required***

533 Final storage
***Required***
***Storage temperature at original site***

☐ 1 Liquid nitrogen ☐ 3 -70°/-80°
☐ 4 -20° ☐ 5 4°
☐ 6 Room temperature ☐ 9 Unknown/refused

529 Specimen stored (type)
***Required***

☐ 2 Serum ☐ 3 Plasma
☐ 25 Mononuclear cells ☐ 95 Other

569 Date of specimen collection ___ Month ___ Day ___ Year
***Required (DMCC states this will be in Specimen Tracking System)***

747 Additive used in blood collection?
***Required***

☐ 1 EDTA ☐ 2 Heparin
☐ 3 Citrate ☐ 4 None
☐ 99 Unknown/refused

751 Original blood sample was collected as:
***Required***

☐ 1 Fasting ☐ 2 Random
☐ 9 Unknown/refused

1248 Was sample collected and processed according to Standard Operating Procedures (SOPs)?
***Required***

☐ 0 No ☐ 1 Yes
☐ Unknown/refused

9

530 Approximate total amount stored ⌊_|__|_⌋
***Required***

531 Approximate total amount stored (Unit)
***Required***

☐ 1 Microliters (mcl)

☐ 2 Milliliters (ml)

☐ 6 Liters (l)

945 Number of freeze-thaws:
***Required***

1267 Number of aliquots sent to NCI Frederick: ⌊_|_|_⌋
***Required***

E.2. DESIRED

## *Group Name: Lung Reference Set*
## *CDEs Desired*

---

799 Height [in inches] (What is your total current height in inches?)
***Desired***

1212 Weight (lbs) at time of diagnosis:
***Desired***

435 Hispanic or Latino (Are you Hispanic or Latino?)
***Desired***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

442 Age first began smoking cigarettes regularly, at least one a day? (How old were you when you began smoking cigarettes regularly, at least one a day?)
***Desired***
***Enter 999 if Unknown/Refused***

445 Average number of cigarettes smoked per day? (During the time you have smoked, on average, how many cigarettes did you smoke per day?)
***Desired***
***Enter 999 if Unknown/Refused***

752 Ever smoke cigars regularly, at least one cigar a day, for a year or more? (Have you ever smoked cigars regularly, at least one a day, for a year or more?)
***Desired***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

1211 Do you now or did you ever smoke a pipe for a year or longer?
***Desired***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

1225 Do you now or did you ever chew tobacco for a year or longer?
***Desired***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

1231 Is there a smoker in participant's household?
***Desired***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

1232 How many years was participant exposed to second hand smoke in the home?

***Desired***

## Occupational History

1229 Has participant been exposed to any of the following known lung carcinogens greater than 8 hours per week for one year? (Check all that apply.)
***Desired***

- ☐ 1 Asbestos
- ☐ 2 Radon
- ☐ 3 Uranium
- ☐ 4 Silica
- ☐ 5 Coal dust
- ☐ 44 None

1230 Exposure to known lung carcinogens (other, specify):
***Desired***

## Medication Use

1233 If you use any illicit drugs, please specify:
***Desired***

## Demographics

1214 What is your living environment?
***Desired***

- ☐ 1 Live in own home
- ☐ 2 Live in assisted living
- ☐ 3 Live in a nursing home
- ☐ 4 Live with child/children
- ☐ 5 Live with friends
- ☐ 7 Other, specify:

1215 What is your living environment (other, specify):
***Desired***

1226 If you live at home, who else lives with you?
***Desired***

1227 If you have children, do they live within an hour's drive?
***Desired***

## Alcohol Consumption

997 Has participant ever had at least one drink of alcohol [beer, liquor, wine, or wine coolers] per month during a twelve-month period? (Have you ever had at least one drink of alcohol [beer, liquor, wine, or wine coolers] per month during a twelve-month period?)
***Desired***

- ☐ 0 No
- ☐ 1 Yes
- ☐ 9 Unknown/refused

782 On average, how many shots of hard liquor or mixed drinks do you drink? Count one shot (1 1/2 ounces) or one mixed drink as one drink.
***Desired***

- ☐ 4 None
- ☐ 5 <1 per week

☐ 6 1-6 per week
☐ 7 1-2 per day
☐ 8 3-5 per day
☐ 9 6+ per day

783 On average, how many glasses of wine do you drink? Count a four-ounce glass of wine as one drink.
***Desired***

☐ 4 None
☐ 5 <1 per week
☐ 6 1-6 per week
☐ 7 1-2 per day
☐ 8 3-5 per day
☐ 9 6+ per day

784 On average, how many glasses/cans of beer do you drink? Count a twelve-ounce can as one beer.
***Desired***

☐ 4 None
☐ 5 <1 per week
☐ 6 1-6 per week
☐ 7 1-2 per day
☐ 8 3-5 per day
☐ 9 6+ per day

## Medical History

1271 Comorbidities:
***Desired***

☐ 1 Myocardial infarction
☐ 2 Congestive heart failure
☐ 3 Peripheral vascular disease
☐ 4 Cerebrovascular disease
☐ 5 Dementia
☐ 6 Chronic pulmonary disease
☐ 7 Connective tissue disease
☐ 8 Ulcer disease
☐ 9 Mild liver disease
☐ 10 Diabetes
☐ 11 Hemiplegia
☐ 12 Moderate or severe renal disease
☐ 13 Diabetes with end organ damage
☐ 14 Moderate or severe liver disease
☐ 15 Metastatic solid tumor
☐ 16 AIDS

1047 Cause of death:
***Desired***

1266 Other lung diseases participant has:
***Desired***

☐ 1 COPD
☐ 2 Asthma
☐ 3 Pulmonary fibrosis
☐ 4 Sarcoidosis
☐ 5 Granulomatosis
☐ 6 Cavitary lesion
☐ 7 Bronchiectasis

1234 Chest X-ray date: ___ ___ ___ (Month Day Year)

***Desired for Set A***

1235 Chest X-ray nodule size (cm): ___
***Desired for Set A***

1236 Chest X-ray nodule location:
***Desired for Set A***

1237 Chest CT date: ___ ___ ___ (Month Day Year)
***Desired for Set A***

1238 Chest CT nodule size (cm): ___
***Desired for Set A***

1239 Chest CT nodule location:
***Desired for Set A***

1240 Reported adenopathy > 1 cm in the mediastinum at time of diagnosis?
***Desired for Set A***

☐ 0 No ☐ 1 Yes
☐ 9 Unknown/refused

1268 PET scan date: ___ ___ ___ (Month Day Year)
***Desired***

1269 PET scan standardized uptake value (SUV) for the lesion:
***Desired***

☐ 1 Less than 2.5 ☐ 2 Greater than or equal to 2.5

1273 PET scan standardized uptake value (SUV) for the liver:
***Desired***

1274 PET scan other hotspots:
***Desired***

☐ 1 Mediastinum ☐ 2 Bone
☐ 3 Brain ☐ 5 Adrenal
☐ 10 Liver ☐ 12 Lung
☐ 44 None

## Lung

1249 Pulmonary function test date (MM/DD/YYYY): ___ ___ ___ (Month Day Year)
***Desired***

1250 FVC liters
***Desired***

1251 FVC % predicted:
***Desired***

1252 FEV1 liters:
***Desired***

1253 FEV1 % predicted:
***Desired***

1254 FEV1/FVC
***Desired***

1255 DLCO liters/minute/mmhg

***Desired***

1256 DLCO % predicted:
***Desired***

1241 Histologic dimensions - height (cm): ___ . ___
***Desired***

1242 Histologic dimensions - width (cm): ___ . ___
***Desired***

1243 Histologic dimensions - depth (cm): ___ . ___
***Desired***

1244 Histologic dimensions - volume (ml): ___ . ___
***Desired***

1245 Distance/pleural margin (mm): ___
***Desired***

1246 Distance/bronchial margin (mm): ___
***Desired***

720 Anatomical site
***Desired***

- ☐ 1 Bladder
- ☐ 2 Bladder peritoneum
- ☐ 3 Bowel peritoneum/serosa
- ☐ 4 Corpus
- ☐ 5 Cervix
- ☐ 6 Cul-de-sac
- ☐ 7 Diaphram
- ☐ 8 Fallopian tube
- ☐ 9 Omentum
- ☐ 10 Paracolic gutter
- ☐ 11 Pelvic wall
- ☐ 12 Peritoneum/uterine serosa
- ☐ 13 Vagina
- ☐ 14 Ovary
- ☐ 15 Mainstem bronchus
- ☐ 16 Carina
- ☐ 17 Upper lobe (UL)
- ☐ 18 Lower lobe (LL)
- ☐ 19 Middle lobe (ML)
- ☐ 20 Ureter
- ☐ 21 Renal pelvis
- ☐ 22 Endometrium
- ☐ 25 Lymph node
- ☐ 95 Other
- ☐ 97 Other, specify:

## Cancer History

1210 Primary cancer treatment(s) received:
***Desired***

- ☐ 1 Chemotherapy
- ☐ 2 Radiation therapy
- ☐ 3 Surgery
- ☐ 9 Unknown/refused

☐ 95 Other

1258 Chemotherapy treatment(s) received: [specify drug(s)]
***Desired***

1259 Chemotherapy start date: (MM/YYYY) Month Day Year
***Desired***

1260 Chemotherapy end date: (MM/YYYY) Month Day Year
***Desired***

1261 Radiation therapy received: (specify site)
***Desired***

1262 Radiation therapy start date: (MM/YYYY) Month Year
***Desired***

1263 Radiation therapy end date: (MM/YYYY) Month Year
***Desired***

1264 Total dose to chest: (cGy) ____.____
***Desired***

1265 Total dose to other organ(s): (cGy) ____.____
***Desired***

1282 Radiation therapy (RT) volume [cubic mm]:
***Desired***

1281 Radiation therapy (RT) complications:
***Desired***

☐ 1 Pneumonitis requiring steroids
☐ 2 Esophagitis requiring IVF
☐ 3 Feeding tube
☐ 4 Hospitalization
☐ 97 Other, specify:

1279 Response to treatment (imaging criteria)
***Desired***

☐ 0 No
☐ 1 Yes

1278 Date of response (MM/DD/YYYY) Month Day Year
***Desired***

1220 Disease status:
***Desired***

☐ 3 Stable Disease
☐ 4 Progression
☐ 5 Recurrence/relapse
☐ 6 No evidence of disease
☐ 9 Unknown/refused

1276 Disease free survival (DFS)?
***Desired***

☐ 0 No
☐ 1 Yes

1280 Overall survival (OS) in days: ____
***Desired***

1221 Recurrence date (MM/YYYY): Month Year

***Desired***

1222 Recurrence site:
***Desired***

- ☐ 1 Bladder
- ☐ 2 Bone
- ☐ 3 Brain
- ☐ 4 Breast
- ☐ 5 Cervix
- ☐ 6 Colon
- ☐ 7 Esophagus
- ☐ 8 Head & neck (mouth, nose, and throat)
- ☐ 9 Kidney
- ☐ 10 Liver
- ☐ 11 Leukemia
- ☐ 12 Lung
- ☐ 13 Lymphoma, including Hodgkins
- ☐ 14 Ovary
- ☐ 15 Pancreas
- ☐ 16 Prostate
- ☐ 17 Rectum
- ☐ 18 Skin (melanoma, no basal or squamous)
- ☐ 19 Stomach
- ☐ 20 Thyroid
- ☐ 21 Uterus
- ☐ 24 Testis
- ☐ 26 Vagina
- ☐ 44 None
- ☐ 95 Other
- ☐ 99 Unknown/refused

1277 Time to local recurrence (TTLR): ___
***Desired***

1223 Progression date (MM/YYYY) ____ ____
***Desired***

1275 Time to progression in days (TTP): ___
***Desired***

1224 Progression site:
***Desired***

- ☐ 1 Bladder
- ☐ 2 Bone
- ☐ 3 Brain
- ☐ 4 Breast
- ☐ 5 Cervix
- ☐ 6 Colon
- ☐ 7 Esophagus
- ☐ 8 Head & neck (mouth, nose, and throat)
- ☐ 9 Kidney
- ☐ 10 Liver
- ☐ 11 Leukemia
- ☐ 12 Lung
- ☐ 13 Lymphoma, including Hodgkins
- ☐ 14 Ovary
- ☐ 15 Pancreas
- ☐ 16 Prostate
- ☐ 17 Rectum
- ☐ 18 Skin (melanoma, no basal or squamous)

- ☐ 19 Stomach
- ☐ 20 Thyroid
- ☐ 21 Uterus
- ☐ 24 Testis
- ☐ 26 Vagina
- ☐ 44 None
- ☐ 95 Other
- ☐ 99 Unknown/refused

## Medical History

1045 Performance status (Which of the following options would you say describes your current performance status?)
***Desired***

- ☐ 0 Fully active, able to carry on all pre-disease performance without restriction
- ☐ 1 Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
- ☐ 2 Ambulatory and capable of all selfcare but unable to carry out any work activities. Up and about more than 50% of waking hours
- ☐ 3 Capable of only limited selfcare, confined to bed or chair more than 50% of waking hours
- ☐ 4 Completely disabled. Cannot carry on any selfcare. Totally confined to bed or chair

1247 Time from blood draw to first freeze: (value) ⌊__⌋
***Desired***

## Appendix F. Material Transfer Agreement form

### Agreement for the Transfer of Materials to the NCI

This Agreement is used for transfers of human biospecimens in research studies to determine the robustness of new cancer biomarkers for possible investigation in Phase II validation studies with the Lung Cancer Proteomics Committee represented by the Special Programs of Research Excellence ("SPORE") and the Early Detection Research Network ("EDRN"), multi-institution networks of investigators funded by the National Cancer Institute ("NCI").

PROVIDER: *[please insert name of institution and scientist providing MATERIAL to the NCI]*

1. The PROVIDER agrees to transfer to the NCI the following MATERIAL, which is the property of the PROVIDER: *[please insert description of samples to be provided]* (hereinafter referred to as "MATERIAL").

2. THIS MATERIAL MAY NOT BE USED IN HUMAN SUBJECTS.

3. The MATERIAL has been collected from human subjects. The NCI will not receive any private identifiable patient information. The MATERIAL has been collected under an IRB approved protocol *[please insert title and number of the IRB approved protocol]*, which includes all necessary informed consents and authorizations which disclose potential redistributions of the MATERIAL in accordance with Section 5 of this Agreement, in accordance with all applicable federal regulations for the protection of human subjects, including, as applicable, 45 CFR Part 46, "Protection of Human Subjects," and the *Standards for Privacy of Individually Identifiable Health Information* set forth in 45 C.F.R. Part 164, under Assurance Number *[please insert number]*. NCI will only transfer MATERIAL and any clinical data, results and raw data relating to the MATERIAL that is stripped of identifiable private information.

4. Any MATERIAL delivered pursuant to this Agreement is understood to be experimental in nature and may have hazardous properties. THE PROVIDER MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED. THERE ARE NO EXPRESS OR IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, OR THAT THE USE OF THE MATERIAL WILL NOT INFRINGE ANY PATENT, COPYRIGHT, TRADEMARK, OR OTHER PROPRIETARY RIGHTS. No indemnification for any loss, claim, damage, or liability is intended or provided by any party under this agreement. Unless prohibited by law, the NCI assumes all liability for claims for damages against it by third parties which may arise from the NCI's use, storage or disposal of the MATERIAL except that, to the extent permitted by law, the PROVIDER shall be liable to the NCI when the damage is caused by the gross negligence or willful misconduct of the PROVIDER.

5. The MATERIAL will be redistributed by the NCI's contractor on behalf of NCI in accordance with criteria established by the EDRN Executive Committee and the NCI Division of Cancer Prevention's Cancer Biomarkers Research Group to qualified investigators and their institutions (RECIPIENTS) who shall first have executed written agreements with the NCI.

6. The MATERIAL is provided at no cost.

7. Any inventions arising from a RECIPIENT's use of the MATERIAL shall be governed by U.S. patent law. RECIPIENTS shall retain title to any patent or other intellectual property rights in inventions made by its employees in the course of conducting research with the MATERIAL. No right, title or interest in any such invention is transferred by virtue of this Agreement.

The PROVIDER and the NCI must sign both copies of this letter and return one signed copy to the PROVIDER. The PROVIDER will then send the MATERIAL.

FOR THE PROVIDER:

Provider Scientist: _____
Provider Organization: _____
_____
_____

Name of Authorized Official: _____
Title of Authorized Official: _____
Address: _____
_____

Signature of Authorized Official: _____
Date: _____

FOR THE NCI:

Name of Authorized Official: _____
Title of Authorized Official: _____
Signature of Authorized Official: _____
Date: _____

## References.

(*1*) Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. Clin Chem 2005;51:102-12.

(*2*) Grizzle WE, Adam BL, Bigbee WL, Conrads TP, Carroll C, Feng Z, et al. Serum protein expression profiling for cancer detection: validation of a SELDI-based approach for prostate cancer. Dis Markers 2003;19:185-95.

(*3*) Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 1999;354:99-105.

(*4*) Swensen SJ, Jett JR, Sloan JA, Midthun DE, Hartman TE, Sykes AM, et al. Screening for lung cancer with low-dose spiral computed tomography. Am J Respir Crit Care Med 2002;165:508-13.

(*5*) Sobue T, Moriyama N, Kaneko M, Kusumoto M, Kobayashi T, Tsuchiya R, et al. Screening for lung cancer with low-dose helical computed tomography: anti-lung cancer association project. J Clin Oncol 2002;20:911-20.

(*6*) Sone S, Li F, Yang ZG, Honda T, Maruyama Y, Takashima S, et al. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. Br J Cancer 2001;84:25-32.

(*7*) Pastorino U, Bellomi M, Landoni C, De Fiori E, Arnaldi P, Picchio M, et al. Early lung-cancer detection with spiral CT and positron emission tomography in heavy smokers: 2-year results. Lancet 2003;362:593-7.

(*8*) Sullivan Pepe M, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054-61.